Working Paper
ENGLISH
ONLY

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)

CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION

STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Helsinki, Finland, 5 to 7 October 2015)

Topic (iv): Access to Statistical Data for Scientific Purposes

Circle of Trust for International Microdata Access

Maurice Brandt*

* Federal Statistical Office, Research Data Centre, Gustav-Stresemann-Ring 11, 65189 Wiesbaden, Germany, e-mail: maurice.brandt@destatis.de

Abstract: Transborder access or usage of international microdata is still a challenge; for researchers and data providers. Researchers often have to deal with the comparability of different national microdata sets across countries and the procedure to get access to those data to combine or analyse them at a central place. Data providers often have to deal with the National Statistical Law which might be more or less precise on microdata access for foreign researchers.
The concept of the "circle of trust for international microdata access" is based on the agreement that each member is accepted according to the same rules and conditions that are approved by all members. In the field of statistical information, this will mainly refer to confidentiality rules and security requirements, but also to competences, legal and technical aspects. According to the level of confidentiality, one may envisage different zones of trust in the circle, whereas the inner circle is the most sensitive and secure.
For security and organisational reasons it is also possible to combine the different ways of access, like a safe centre and remote access. Usually the microdata are physically in a safe centre, but in combination with remote access it is also possible that the data are accessed from a safe centre but the secure server where the data are stored is located somewhere else. This might be a good approach if the data producing authority does not send out any data and the access is not allowed to be granted at the researcher's institution. In this case the researcher can visit an accredited safe centre that is connected remotely to the data at another countries server. Based on the different modes of access combined with different levels of data confidentiality several zones of trust can be implemented. There are already some good examples in the EU that could provide the basis for an international circle of trust for sharing microdata services.

1 Introduction

The microdata access to official statistics on national level is basically available in many countries. The access to EU microdata is possible in the framework of the European Statistical System (ESS), organized on a legal basis with the regulation (EU) No 557/2013, which enables the National Statistical Institutes and Eurostat to operate on a well-defined procedure. Many efforts have been made to improve the infrastructure for researchers by investigating in a decentralised system for microdata access in Europe. This was started with the ESSnet-Projects "Decentralised Access to EU Microdata Sets" and "Decentralised and Remote Access to Confidential Data in the ESS" (Brandt 2013). However, this concerns only the already prepared and harmonised EU microdata that are available at Eurostat.

For an individual access to many other national microdata sets across countries the challenges are even bigger. Transborder access or usage of international microdata is still at the beginning for several reasons. Absence of appropriate laws, technical infrastructure, English metadata and harmonised standards are just a few of them. The European FP7-project "Data without Boundaries" has initiated the process to collaborate with partners of Official Statistics, Social Science Data Archives, universities and research institutes to improve the international access to national microdata sets. This is a quite a challenge, because there is no unified legal umbrella, organizational framework or technical infrastructure yet for this purpose. But there might be an approach to change this situation.

For sharing microdata services, i.e. exchange of microdata or providing access to confidential microdata to third parties, a concept of basic requirements is necessary.

According to the level of data confidentiality there are also different grades of risk assessments. This means that for original data (without direct identifiers) exists a very high risk of disclosure and for public use files almost no risk at all.

The concept of the "circle of trust" for international microdata access respects the different levels of data confidentiality and therefore the different levels of security requirements that are necessary to access those data.

Many aspects of international microdata access is regulated by the national statistical act. If the statistical act prohibits granting access to foreign third parties, there is no trust needed. If the statistical act allows giving access, there might be no trust needed as well. But there is also the situation that access for another country or a third party is not mentioned in the national law, what means that it is not forbidden from a legal point of view. There are two interpretations of the law in this case. Some countries treat everything what is not explicitly forbidden as allowed and some countries treat everything what is not explicitly allowed as forbidden. So, even when giving access to third parties is not forbidden, some of the National Statistical Authorities still don't want to give access as long as they are not obliged by law. From this perspective the "circle of trust concept" is needed for three reasons.

The first one is that a "circle of trust" can bridge the gap for those countries, whose law does not mention anything at all. For them it would be helpful to have a basic concept of requirements so that they are reassured that their data are accessed in an organized and secure manner comparable with the security requirements in their own

country. The second reason is that a circle of trust can even reassure those countries that basically allow access to foreign third parties. A law can be broken, but it is possible to prevent a breach of confidentiality with an organizational or technical solution and a reasonable amount of effort. Their data will be accessed at least under the same security standards like in their own country. The third reason is that those countries, that do not allow microdata access to foreign third parties at all, need to rethink their position, if the data are accessed in a very secure and modern way at least comparable with their own standards, there is actually no practical reason to refuse microdata access in the long run. There is always the opportunity to deliver anonymised datasets for the beginning. The report of the OECD "Expert Group for International Collaboration on Microdata Access" gives a detailed overview about the measures that can be made to improve the international microdata access. There are maturity indicators described with recommendations to move from low to medium level and from medium to high level with an embedded infrastructure.

2 The concept

The concept of "circle of trust" is based on the agreement that each member is accepted according to the same rules and conditions that are approved by all members. In case of statistics, this will mainly refer to confidentiality rules and security requirements but also to competence and legal aspects. This makes it possible to create a group / membership of trust. According to the level of confidentiality, one may envisage different zones of trust in the circle, whereas the inner circle is the most sensitive and secure.

For security or organisational reasons it is also possible to combine the different ways of access, like a safe centre and remote access. Usually the microdata are physically in a safe centre, but in combination with remote access it is also possible that the data are accessed from a safe centre but the secure server where the data are stored is located somewhere else. This might be a good approach if the data producing authority does not send out any data and the access is not allowed to be granted at the researcher's institution. In this case the researcher can visit a accredited safe centre that is connected remotely to the data at another countries server. Based on the different modes of access combined with a different level of data confidentiality several zones of trust can be implemented.

Figure 1 shows how the different zones can look like and what kind of data access is adequate for each zone. The dimension goes from highly confidential microdata in the inner circle to less confidential in the outer circle. Outside the circle means non-confidential, only public use files or open data will be provided to everybody, i.e. companies and private persons.

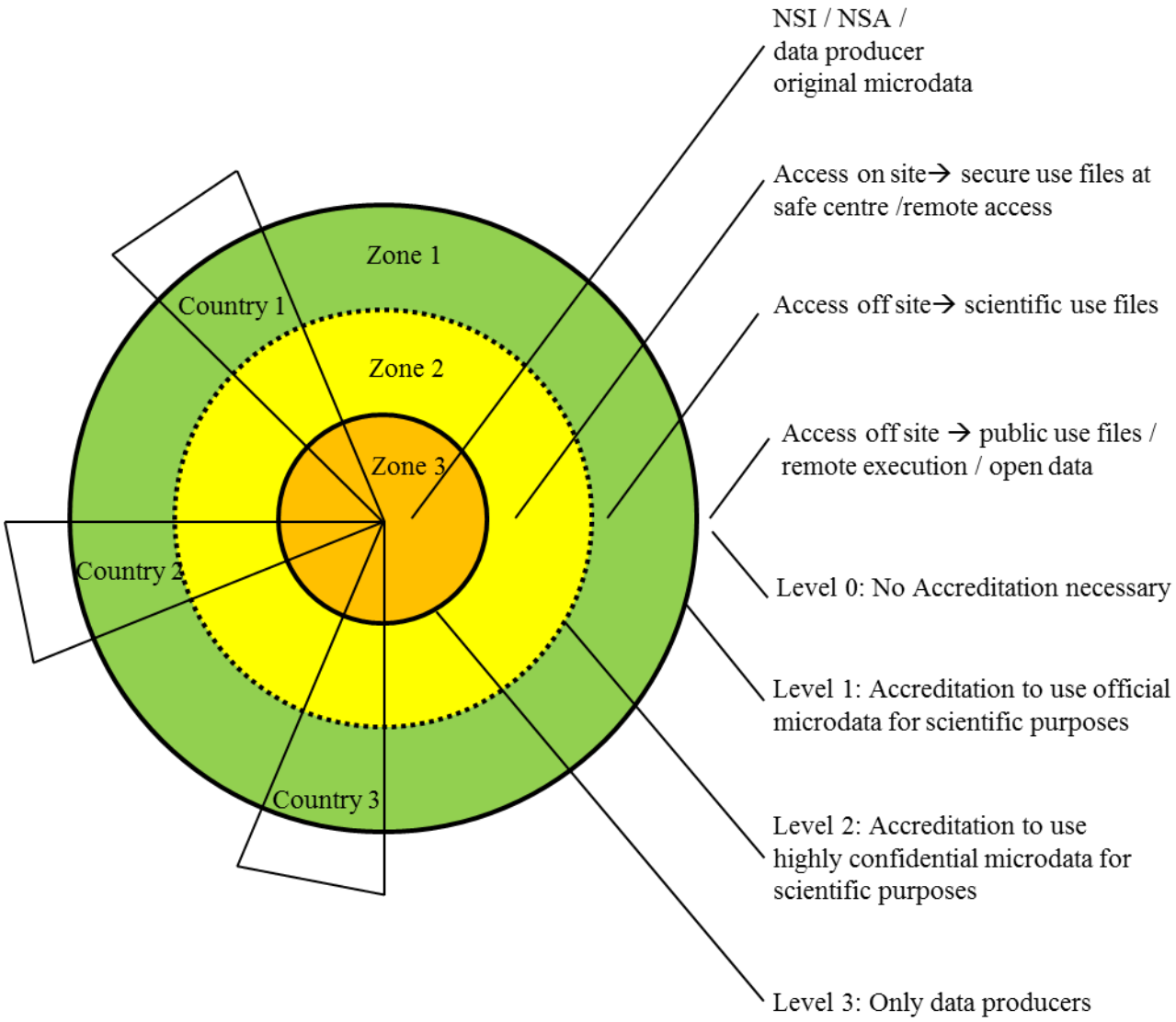

Figure 1: levels of **"**circle of trust" with different ways of access and different kind of microdata

3 The elements of trust

This concept does not mean that data producers only have to trust that the sensitive information of their data will be treated properly. The trust is generated according to specific rules, standards, protocols and guidelines that each participating party has to fulfil. These measures can be:

- glossary of terms and definitions for a common understanding

- share best practices

- collection and documentation of rules and protocols for transparency

- cooperation agreement
- harmonised templates and contracts for microdata access
- national guidelines for the treatment of international or foreign microdata requests
- catalogue of rules to check which institution is approved to access microdata
- accreditation criteria for scientific institutions
- security concept and accreditation guidelines for safe centres
- list of security and user demands for a remote access system
- anonymisation concept for scientific use files
- rules and protocols for the transmission of microdata
- data protection guidelines
- common view on IT security standards
- guidelines for statistical disclosure methods and output checking
- administrative sanctions / penal sanctions (if applicable)
- common understanding of responsibilities and similarities
- legal framework

The preconditions can be set up for the institution itself and for the details of the technology that is providing the access. The minimum requirement for most data producers is that data which are taken from another body is treated at least under the same or higher level of confidentiality as in their own premises.

The zones itself represent the microdata with the different levels of access and confidentiality and the institutions that are accredited to access the microdata. The lines between the circles represent the different levels of confidentiality and therefore the requirements that an institution has to fulfil to access the data in the respective zone. To enter the circle at all to zone 1 a minimum of criteria has to be fulfilled and level 1 needs to be passed. Some countries may not differentiate between zone 1 and 2 (dashed line) and there is only a binary decision. Whether an institution is accredited as scientific or not and if they are considered as scientific, they can access all kind of data. The triangles represent the countries whereas a country can host the data producer and the data requesting institution as well. The apex of the triangle in the green zone 3 is only for the data producing authority and can usually not be accessed by an external institution.

As a first trust building mechanism, an agreement on the rules and standards for joining the “circle of trust” will be helpful and also a common understanding on which parties are expected to be in the inner or outer circle.

It would be advantageous, if a research institution that is accredited by one country for zone 2, would be automatically accredited in another country for zone 2. This means the institution can use secure use files of all countries in the ring of zone 2. Before that an investigation is necessary if all countries have the same understanding of the confidentiality levels of the different zones, it need to be checked if they even

have that kind of distinction. Also a harmonised accreditation system need to be applied that has been agreed by all countries. There are still several open questions to be answered first and such an ambitious system of international microdata access should be future proof in terms of technology. In case of consideration of the new technologies like cloud computing, the data need to be stored in a central place to prepare an international dataset out of single countries. This detail will remain very sensitive because it means a physically transmission of the data.

4 Outlook

In terms of trust, it is very important to consider what already exists and to build on that. This can be an already existing infrastructure or a best practice system in a single country. The data producing authorities are also depending on the trust of the respondents, what means that a high quality production system for official statistics can only be guaranteed as long as this trust can be maintained. Trust is of course not a one way street that goes only in one direction. It needs to be proved over many years and one single negative experience can destroy all the investments and efforts built in the past.

The responsibility for that lays with the data producers, because respondents mostly do not care who is at the end responsible for the breach of confidentiality. Their only possibility of protection is to refuse a survey, and in case they are obliged by law, they might give incorrect answers. This scenario would decrease the quality of official statistics. It would also affect not only the data producers, but also all the users of the statistics and microdata as well. The need for microdata is absolutely comprehensible, but researchers are also interested in good quality of the data. Therefore the producers should be still in charge of the admissibility criteria and the decision of the parties that are accessing their data.

This model of sharing microdata services is not a system that can be established in a short run. It is more a development of gaining trust, experience and competences in this field and the need to prove that this way of microdata access is secure. It might be that trust is not enough and therefore guarantees are needed. The suggestion would be to start small with a few countries to prove the feasibility, security and the concept under realistic conditions. A good example for a good cooperation is the European Statistical System (ESS). It provides EU institutions, Member States and the public with reliable information about the society, economy, environment and development in the European Union. There is also an established access to the EU microdata for the international research community. This process is quite elaborative, because every request needs to be coordinated with 28 Member States. The procedure is very well organized so that research institutions, once they are accredited, can get access in a relatively short time. This could be a starting point or a raw model for other initiatives and projects that are striving for a better international research infrastructure.

References

Brandt, Maurice (2013): Improvement of access to European microdata, Joint UNECE/Eurostat work session on statistical data confidentiality, Ottawa.

DwB project website: http://www.dwbproject.org/

Commission Regulation (EU) No 557/2013 of 17 June 2013 implementing Regulation (EC) No 223/2009 of the European Parliament and of the Council on European Statistics as regards access to confidential data for scientific purposes and repealing Commission Regulation (EC) No 831/2002 Text with EEA relevance http://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32013R0557

ESSnet "*Decentralised Access to EU-Microdata Sets*" *(DA):* http://www.cros-portal.eu/content/da-finished

ESSnet *"Decentralised and Remote Access to Confidential Data in the ESS (DARA):* http://www.cros-portal.eu/content/decentralised-and-remote-access-confidential-data-ess-dara

Eurostat microdata: (http://ec.europa.eu/eurostat/web/microdata/overview)

OECD Expert Group for International Collaboration on Microdata Access, Final Report: http://www.oecd.org/std/microdata-access-final-report-OECD-2014.pdf